Guía Criterios de Disociación de Datos Personales.

Resumen

Versión 2.0 – AÑO 2017

La Unidad Reguladora y de Control de Datos Personales, por Resolución Nº 68/017, de 26 de abril de 2017, aprobó el documento Criterios de Disociación de Datos Personales conforme lo dispuesto por el Decreto N° 54/017, de 20 de febrero de 2017, reglamentario del artículo 82 de la Ley N° 19.355, de 19 de diciembre de 2015, en el que se establece que las Entidades Públicas, sujetos obligados por la Ley N° 18.381,de 17 de octubre de 2008, deberán proceder a la publicación de la información contenida en los artículos 5° de la Ley y 38 y 40 del Decreto N° 232/010, 2 de agosto de 2010, en formato de dato abierto.

En este e-book se encontrarán los diversos criterios de no identificación de aquellos datos personales que se encuentran en la información que se deba publicar como dato abierto, como ser: seudonimización, disociación, anonimización entre otros.

Por lo tanto con estos criterios se pretende, que toda persona que utilice estas técnicas incorpore los lineamientos dados en este e-book, para disminuir al mínimo la posibilidad de reidentificar al titular del dato que se maneje, teniendo en cuenta que esta es una actividad dinámica que varía por la constante aparición de nuevos mecanismos de reidentificación.

Capítulo 1. Introducción

Con el constante desarrollo y evolución de la tecnología, los diferentes grupos sociales se han transformado en grandes creadores y acopiadores de datos. Estos al acumularse forman un conjunto que se convierte en información.

La generación de la información se debe a que dichas entidades al cumplir con sus actividades, realizar sus actividades, utilizar diversos dispositivos electrónicos y su software (computadoras personales, tablets, celulares, programas, aplicaciones, entre otros) la intercambian y almacenan obteniendo un activo determinante.

Cada uno de los cometidos desarrolladas por el Estado o por las empresas y particulares, debe realizarse dentro del marco normativo vigente de protección de datos personales, para garantizar los derechos de acceso, rectificación, actualización y supresión.

El tratamiento de los datos que se efectúa es diverso según quién realiza la actividad (Estado, empresa, personas físicas), primando en el sector privado la autonomía de la voluntad y en el sector público la actuación acorde a Derecho en atención, entre otros, al principio de especialidad, por tanto estas distinciones serán reflejadas a lo largo de estos criterios.

En la actualidad ha cobrado especial interés la utilización de la información que se encuentra en poder del Estado, la que puede ser utilizada en beneficio de la sociedad mediante las herramientas que brinda la apertura de datos. Para poder realizar esta actividad es necesario tomar algunos recaudos y evitar la vulneración de los derechos de terceros implicados.

Estos recaudos se concretan, con el cumplimiento de la normativa vigente en el tema, y también con las técnicas de anonimización o disociación de los datos, para evitar los riesgos que pueden llevar a la vulneración de la identidad de los titulares y su reidentificación.

Al mismo tiempo, es posible que los datos en los que se han empleado las técnicas de anonimización o disociación, aún gocen de una protección dada por normativa que no sea la de protección de datos, como por ejemplo la Ley Nº 18.381, de 17 de octubre de 2008, de Acceso a la Información Pública.

En estos criterios se pretende, que toda persona que utilice las técnicas de anonimización o de disociación del dato del titular, incorpore sus lineamientos dados, para disminuir al mínimo la posibilidad de reidentificar al titular del dato que se maneje, teniendo en cuenta que esta es una actividad dinámica que varía por la constante aparición de nuevos mecanismos de reidentificación.

Además, se vuelve relevante la adopción de un plan de trabajo que permita realizar el proceso de anonimización en forma ordenada, para lo que es necesario transitar por una etapa de preanonimización, una de anonimización y una de control.

Capítulo 2. Definiciones

Las definiciones recogidas en la Ley N° 18.331 y en el decreto 414/009, muchas veces pueden resultar un exceso de información pero en este caso, en el que el tema es nuevo y que es necesario unificar conceptos y conocimientos, resulta imperioso incorporarlas para ser más ilustrativo su contenido:

- **Dato personal:** es toda información de cualquier tipo, referida a personas físicas o jurídicas, determinadas o determinables, como: cualquier información numérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo que a ellas se refiera.
- **Datos sensibles:** son datos personales que revelen origen racial y étnico, preferencias políticas, convicciones religiosas o morales, afiliación sindical e informaciones referentes a la salud o a la vida sexual.
- **Titular del dato:** persona cuyos datos sean objeto de un tratamiento incluido dentro del ámbito de acción de la presente ley.
- **Tratamiento de los datos:** operaciones y procedimientos sistemáticos, de carácter automatizado o no, que permitan el procesamiento de datos personales, así como también su cesión a terceros a través de comunicaciones, consultas, interconexiones o transferencias.
- **Usuario de datos:** toda persona, pública o privada, que realice a su arbitrio el tratamiento de datos, ya sea en una base de datos propia o a través de conexión con los mismos.
- **Disociación:** es todo tratamiento de datos personales de manera que la información obtenida no pueda vincularse a persona determinada o determinable.
- **Anonimización:** es el resultado de un tratamiento de los datos personales realizado para evitar de forma irreversible su identificación (Dictamen N° 05/2014, adoptado por el Grupo de Trabajo del art. 29 de la Directiva 95/46/CE).
- **Reidentificación:** es volver a identificar o identificarse un titular del dato luego de un tratamiento de anonimización o disociación de sus datos personales.
- **Seudonimización:** reduce el vínculo de un conjunto de datos con la identidad original del interesado , o tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado en particular sin recurrir a información adicional, siempre que dicha información adicional se mantenga separada y sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos no se atribuyan a personas identificadas o identificables (Reglamento (UE) Nº 679 /2016 del Parlamento Europeo y del Consejo relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos -Reglamento General de Protección de Datos).
- **Inicialización:** proceso de sustitución de los nombres y/o números de las personas por letras.
- **Cifrado:** dato que se encuentra escrito con letras, símbolos o números que solo pueden comprenderse si se dispone de la clave necesaria para descifrarlos.
- **Principio de Minimización:** los datos para su tratamiento deben ser limitados a lo necesario en relación con los fines para los que son tratados (Dictamen N° 05/2014). Es el equivalente en nuestro Derecho al principio de proporcionalidad derivado del de veracidad (Art. 7° de la Ley N° 18.331).
- **Principio de Divisibilidad:** Si un documento contiene información que pueda ser conocida e información que debe denegarse en virtud de causa legal, se dará acceso a la primera y no a la segunda (Art. 7° del Decreto N°232/010).

Capítulo 3. Etapas del proceso de anonimización.

El responsable de la base de datos que comience el proceso de apertura de datos debe documentar, las diferentes etapas del proceso de anonimización para obtener un control y resultado óptimos y evitar el posible riesgo de reidentificación.

Las etapas son tres y se denominan:

- preanonimización
- anonimización
- control

3.1. Preanonimización

En la etapa de preanonimización se diseña el proyecto de anonimización, en el que se deberán identificar, con claridad las variables, los identificadores directos e indirectos, los datos confidenciales, cuál o cuáles serán las técnicas adecuadas de anonimización, según el conjunto de datos de que se trate, el riesgo de reidentificación asociado, finalizando con la ejecución del proyecto.

Por variables nos referimos entre otras a: nombres, números de teléfono, números de cuentas bancarias, correos electrónicos, direcciones IP, fotografías e imágenes similares, datos biométricos, dirección, domicilio, matrículas de vehículos, celulares y sus números de serie o cualquier otro número de identificación.

Todo ello deberá permitir la ruptura de los eslabones de la relación de identificación-información.

En la preanonimización se deberán definir los perfiles necesarios de los actores (responsables de la tarea, técnicos informáticos, usuarios), para llevar a cabo el proceso de anonimización y la responsabilidad de cada uno de ellos.

Concomitantemente, se debe definir el plan de contingencia para el caso de detectarse un riesgo de reidentificación una vez publicado los datos, los mecanismos ágiles de respuesta para salvaguardar la identidad del titular del dato.
Al mismo tiempo, no todos los actores del proceso tienen por qué acceder a toda la información en todo momento, por lo que es de suma importancia que el responsable de la tarea proporcione solamente la mínima información necesaria a ser utilizada y en consecuencia anonimizada.

Los responsables deben:

a) Clasificar las variables, por ejemplo variable de identificación geográfica, de identificación directa de personas o empresas, de carácter sensible o confidencial, sin restricción para el acceso público.

b) Identificar la información que será incluida en el proceso de anonimización, ya que no toda la información que posee el organismo puede publicarse en formato abierto.

c) Minimizar la cantidad de información personal que exista dentro del conjunto de datos. Si surgen datos sensibles requieren un mayor cuidado para evitar la trazabilidad inversa y definir la necesidad real de utilizar esa información.

Una vez que se tiene elaborado el proyecto y se ejecuta es necesario que se realice un control en forma periódica.

3.2 Anonimización

Cuando al definirse el círculo de protección, que el derecho a la protección de datos plantea, corresponde tener en cuenta los datos personales.

La solución para que se cumpla con la normativa vigente aplicable al tema resulta ser: anonimizar, disociar los datos de manera tal que no se permita identificación alguna. Sin embargo, ello no es tan sencillo y en la práctica suele incurrirse en equivocaciones que hacen dificultosa la tarea mas no imposible.

La anonimización posee diversas características que deben ser tomadas en cuenta cada vez que se va a materializar, para que los datos personales puedan ser utilizados en forma abierta:

a) no puede establecerse vínculo alguno entre el dato y su titular sin un esfuerzo desproporcionado,

b) no puede ser reversible, es decir que es el resultado de un tratamiento de datos personales realizado para impedir que se vuelva atrás con lo efectuado y se identifique al interesado,

c) que en la práctica sea equivalente al de un borrado permanente,

d) que lleve implícito un factor de riesgo que se debe tener en cuenta al valorar las técnicas de anonimización, además de considerarse la gravedad y probabilidad del riesgo en sí mismo.

Desde el punto de vista de la protección de datos personales en la anonimización de datos, debe considerarse que si bien, en principio es favorable pueden aparecer riesgos tales como:

a) permanencia de datos que permiten reidentificar al titular del dato personal,

b) posibilidad de reidentificar mediante inferencias o por vinculación o relación con otros paquetes o conjuntos de datos personales,

c) confusión entre lo que es seudonimización y anonimización,

d) confusión entre lo que es anonimización y disociación.

Si nos detenemos a ver los puntos c) y d) anteriores son muy sutiles las diferencias entre los términos.

Por lo tanto, la anonimización se usa cuando los datos personales se han recogido y tratado de acuerdo con la legislación vigente sobre su conservación, en un formado identificable. Asimismo su resultado debe ser, de acuerdo con el estado actual de la tecnología, equivalente al borrado, es decir se debe garantizar que es imposible recuperar los datos personales del titular.

Siguiendo el plan elaborado por el responsable del proceso de anonimización durante la etapa de preanonimización, los técnicos deberán aplicar las técnicas seleccionadas, los algoritmos necesarios, realizar pruebas de calidad y entregar el resultado al responsable para su aprobación.

El objetivo final de la anonimización es proveer los datos desagregados para que el público en general pueda utilizarlos, sin generar conflictos con los titulares de los datos.

3.3 Control

La tercera etapa del proceso de anonimización implica la realización de controles periódicos por parte de los técnicos en virtud de la aparición de las nuevas tecnologías y métodos para prevenir y evitar los posibles riesgos de reidentificación.

Verificados estos controles, una respuesta rápida es dar de baja el conjunto de datos que causa la reidentificación, para que no sea levantado por los buscadores de Internet y que no sea necesario aplicar las teorías del derecho al olvido, con lo que se evitará, una posible reparación de los daños y perjuicios al titular del dato.

Por su parte, el titular del dato en caso de sentirse identificado por un conjunto de datos, debe comunicar tal extremo al responsable y este tomará las medidas necesarias, a efectos de aplicar debidamente el derecho a la protección de datos, por lo cual es importante poseer mecanismos ágiles de respuesta por los organismos y las empresas como se mencionó.

Capítulo 4. Conjunto de técnicas de anonimización

Luego de realizar una descripción de los conceptos y el marco normativo aplicable, se detallan a continuación los conjuntos de técnicas más utilizadas para efectuar la anonimización.

Estas deberán acompasar la evolución de las tecnologías que surjan y que puedan ser utilizadas para reidentificar a los titulares de los datos según el principio de neutralidad tecnológica.

Los conjuntos de técnicas de anonimización se dividen en *generales y particulares, y dentro de los generales la aleatorización y la generalización.*

La aleatorización se subdivide en *adición de ruido, permutación, y privacidad diferencial; la generalización en agregación, anonimato k, diversidad l y proximidad t.*

4.1. Aleatorización

La aleatorización es un conjunto de técnicas que modifican la veracidad de los datos con el fin de eliminar el vínculo existente entre ellos y su titular. Si los datos se vuelven lo suficientemente ambiguos, no se podrá identificar a una persona concreta.

Este conjunto de técnicas por sí sola no reduce la particularidad de cada uno de los registros, puesto que estos pueden obtenerse a partir de un único interesado. Puede proteger contra ataques o riesgos de inferencia que se basan en información deducida lógicamente a partir de piezas aparentemente inconexas.

Dentro de la aleatorización encontramos la adición de ruido, permutación, privacidad diferencial.

4.1.1. Adición de ruido

La técnica de adición de ruido es la modificación de los atributos del conjunto de datos para que sean menos exactos, conservando no obstante su distribución general.

Si trata un conjunto de datos, cualquier observador supone que los valores son exactos, pero esto solo es cierto hasta determinado punto.

Un ejemplo de ello es cuando se observa la altura de una persona, esta se mide hasta el centímetro más próximo, es decir 1 metro 80, pero el conjunto de datos anonimizados puede englobar valores con una exactitud de + 10 cms., es decir + 1,70 cm. a 1,90 cm.

Si se utiliza esta técnica de manera competente, un tercero no podrá identificar a una persona ni tampoco debería ser capaz de restaurar los datos o de averiguar cómo se han modificado.

Habitualmente, la adición de ruido debe combinarse con otras técnicas de anonimización, como la eliminación de atributos obvios y de cuasi identificadores. El nivel de ruido depende de la cantidad y el tipo de información que se requiera, así como del impacto que tenga la revelación de los atributos protegidos en la privacidad de las personas.

En esta técnica pueden surgir errores, entre los que se encuentran los siguientes:

a) Añadir ruido inconsistente: Si el ruido está fuera de escala y no respeta la lógica entre los atributos de un conjunto de datos, un atacante que acceda a la base de datos podrá filtrar el ruido y, en algunos casos, recuperar las entradas que faltan. Es más, si existen pocos elementos en el conjunto de datos, persistirá la posibilidad de vincular las entradas de datos con ruido con una fuente externa.

b) Pensar que la adición de ruido es una medida suficiente: La adición de ruido es una medida complementaria que hace más difícil que un atacante obtenga los datos personales. A no ser que el ruido sea mayor que la información contenida en el conjunto de datos, no se debe pensar que la adición de ruido es una solución completa para la anonimización.

En determinados casos se producen defectos en vez de errores, tal es el caso, por ejemplo, de la realización de una reidentificación que se lleva a cabo en una base de datos de clientes del proveedor de contenidos de videos.

Los técnicos analizaron las propiedades que tiene la base de datos y la anonimizaron. La empresa la hizo pública, teniendo en cuenta la normativa de protección de datos.

Para ello lo que procedió a hacer fue eliminar todo tipo de información que pudiera identificar al cliente, excepto las valoraciones y las fechas. Se añadió ruido a las valoraciones mejorándolas o empeorándolas ligeramente.

A pesar de ello, se descubrió que se podía identificar de manera unívoca el 99 % de los registros de usuarios en el conjunto de datos usando 8 valoraciones y fechas con errores de 14 días a modo de criterio de selección. Aun rebajando los criterios de selección a 2 valoraciones y un error de 3 días, se podía identificar al 68 % de los usuarios.

No solo son errores y defectos los que se encuentran en la aleatorización adición de ruido, también existen garantías como las que siguen a continuación:

1. Se pueden singularizar los registros de una persona (quizás de manera no identificable), aunque sean menos fiables.
2. Se pueden vincular los registros de una misma persona, pero estos son menos fiables, por lo cual se puede vincular un registro real con uno añadido artificialmente (es decir, vincularlo con el ruido). En algunos casos, una atribución incorrecta puede exponer al interesado a un nivel de riesgo significativo, incluso mayor que en el caso de una atribución correcta.
3. Se pueden llevar a cabo ataques por inferencia, pero la tasa de éxito será menor, además, no se descartan falsos positivos (o falsos negativos).

4.1.2. Permutación

La técnica de permutación consiste en mezclar los valores de los atributos en una tabla para que algunos de ellos puedan vincularse artificialmente a distintos interesados.

Esta es una estrategia útil en el caso de que sea importante conservar la distribución exacta de cada atributo en el conjunto de datos. La permutación podrá considerarse como una forma de adición de ruido.

En la forma clásica de adición de ruido, los atributos se sustituyen por valores aleatorizados. Generar un ruido consistente puede ser una tarea difícil, aparte de que, si la modificación de los valores de los atributos es mínima, puede que no se obtenga el grado de privacidad deseado.

Con las técnicas de permutación, se intercambian los valores contenidos en el conjunto de datos, trasladándolos de un registro a otro. Esta permuta de datos garantiza que el rango y la distribución de valores sean idénticos, no así las correlaciones entre los valores y las personas.

Si dos o más atributos tienen una relación lógica o una correlación estadística y se permutan independientemente del resto, dicha relación quedará destruida. Por consiguiente, sería importante permutar un conjunto de atributos que estén relacionados entre sí a fin de no romper la relación lógica.

En caso contrario, un atacante podría identificar los atributos permutados y revertir la permutación. Por ejemplo, imaginemos el siguiente subconjunto de atributos en un conjunto de datos médicos: razones para la hospitalización, síntomas y servicio hospitalario responsable. En la mayoría de los casos, existirá una estrecha relación lógica entre los valores, de modo que si se llevara a cabo la permutación en uno solo de estos valores, esta técnica sería detectada e incluso podría revertirse. Al igual que ocurre con la adición de ruido, la permutación por sí sola no permite obtener la anonimización, por lo que siempre debe combinarse con el procedimiento de eliminación de atributos obvios o cuasi identificadores.

4.1.3. Privacidad diferencial

La privacidad diferencial a pesar de pertenecer a las técnicas de anonimización por aleatorización adopta una orientación diferente, ya que esta puede realizarse cuando el responsable del tratamiento de datos genera vistas anonimizadas de un conjunto de datos, al mismo tiempo que almacena una copia de los originales.

A su vez, esta indica al responsable del tratamiento cuánto ruido debe añadir, y en qué forma, para obtener las garantías de privacidad necesarias. En este contexto, es especialmente importante una supervisión continua (como mínimo en cada oportunidad de consulta) para evaluar cualquier posibilidad de identificación de una persona en el conjunto de resultados de las consultas. Sin embargo, conviene aclarar que las técnicas de privacidad diferencial no modifican los datos originales. Mientras se conserven los datos originales, el responsable del tratamiento es capaz de identificar a los titulares de los datos a partir de los resultados de las consultas de privacidad diferencial mediante el conjunto de los medios que pueden ser razonablemente utilizados. Estos resultados también deben considerarse como datos personales y se les aplica la normativa vigente en la materia.

Una de las ventajas de la privacidad diferencial consiste en el hecho de que los conjuntos de datos se entregan a terceros autorizados como respuesta a una consulta concreta y no simplemente como consecuencia de la publicación de un único conjunto de datos.

Desde el punto de vista de la protección de datos, la mayor dificultad que existe es poseer la capacidad de generar la cantidad adecuada de ruido, es necesario hacer bastante ruido, ya que es un error frecuente no hacerlo, el que se añade a las respuestas verdaderas a fin de proteger la privacidad de las personas y, al mismo tiempo, preservar la utilidad de las respuestas difundidas.

Además, conviene tener el cuidado de no caer en el error de pensar que los datos son anónimos para el tercero cuando el responsable del tratamiento todavía puede identificar al interesado en la base de datos original mediante el conjunto de medios que pueden ser razonablemente utilizados.

4.2. Generalización

La generalización es la segunda familia de técnicas de anonimización. Este enfoque generaliza o diluye los atributos de los interesados modificando las respectivas escalas u órdenes de magnitud (por ejemplo, sustituyendo una ciudad por una región, o una semana por un mes). Aunque la generalización pueda ser efectiva para descartar la singularización, no permite obtener una anonimización eficaz en todos los casos; en concreto, es necesario aplicar enfoques cuantitativos específicos y complejos para impedir la vinculabilidad y la inferencia.

4.2.1. Agregación y Anonimato k

Las técnicas de agregación y anonimato k tienen el objetivo de impedir que un interesado sea singularizado cuando se le agrupa junto con, al menos, un número k de personas. Ejemplo de ello es, cuando se toma un atributo que equivale a la edad de los funcionarios, formando grupos de intervalos de valores, es decir entre 30 a 40 años, entre 40 y 50, haciendo franjas.

Estos métodos son aplicables cuando la correlación de valores puntuales de atributos puede crear cuasi identificadores.

La carencia principal del modelo de anonimato k es que no impide los ataques por inferencia.

En el ejemplo de anonimato k que se expone a continuación, se trata de una base de datos, (tabla de Excel), con “n” filas y “m” columnas, en la que cada fila representa un registro relacionando con una persona concreta, de un departamento del país que es paciente de un hospital ficticio en Uruguay. Los valores en las columnas describen las características atributos que se asocian con todos los miembros de esa población.

La tabla siguiente es una base de datos no anonimizada que incluye los registros de pacientes mencionados:

Nombre	Edad	Género	Departamento	Religión	Patología
Juana Rivero	19	F	Artigas	Católica	Esclerosis
María González	23	F	Rivera	Hindú	Infección
Ricardo Pérez	43	M	Cerro Largo	Mormón	Ninguna
Anibal Rojas	54	M	Lavalleja	Católica	Renal
Manuel Quinte	59	M	Montevideo	Católica	Cáncer
Jesus Sosa	29	M	Maldonado	Musulmán	Cáncer
Lorena Quiroz	33	F	Rocha	Budismo	Renal
Alejandra Rodo	51	F	Colonia	Católica	Esclerosis
Natalia Pacheco	45	F	Soriano	Hindú	Infección
Javier Medina	34	M	Soriano	Mormón	Renal
Angela Torres	18	F	Artigas	Musulmán	Listeria
Lucas Iglesias	58	M	Tacuarembó	Budismo	Listeria

Tabla 1. Datos no anonimizados con inclusión de los registros de pacientes.

La Tabla 1 consta de 6 atributos y 12 registros. Para obtener un anonimato k existen dos métodos comunes que se pueden utilizar, cada uno con sus riesgos y errores asociados:

a) Supresión: En este método, todos o algunos valores de los atributos son reemplazados por un asterisco “*”, de igual forma los de las columnas. En la tabla anonimizada inferior, todos los valores en el atributo “Nombre” y en el atributo “religión” han sido reemplazados por un “*”.

b) Generalización: En este método, los valores individuales de atributos son reemplazados por una categoría más amplia. Por ejemplo, el valor “19” del atributo “Edad” puede ser reemplazado por “≤ 20”, el valor “23” por “20 < Edad ≤ 30”, etc.

La siguiente tabla muestra la base de datos anonimizada con la aplicación de los métodos descriptos:

Nombre	Edad	Género	Departamento	Religión	Patología
*	Edad≤20	F	Artigas	*	Esclerosis
*	20<Edad≤30	F	Rivera	*	Infección
*	40<Edad≤50	M	Cerro Largo	*	Ninguna
*	50<Edad≤60	M	Lavalleja	*	Renal
*	50<Edad≤60	M	Montevideo	*	Cáncer
*	20<Edad≤30	M	Maldonado	*	Cáncer
*	30<Edad≤40	F	Rocha	*	Renal
*	50<Edad≤60	F	Colonia	*	Esclerosis
*	40<Edad≤50	F	Soriano	*	Infección
*	30<Edad≤40	M	Soriano	*	Renal
*	Edad≤20	F	Artigas	*	Listeria
*	50<Edad≤60	M	Tacuarembó	*	Listeria

Tabla 2. Datos anonimizados con inclusión de los registros de pacientes.

Los datos relativos a los atributos “Edad”, “Género” y “Departamento”, son anónimos, ya que si se combinan cualquiera de ellos se encontrarán al menos 2 filas con idénticos atributos.

Por ejemplo: la fila 1 con la fila 11, en la que se encuentran dos registros con edad < 20, ambas femeninas.

La combinación de los atributos se denomina cuasi identificadores por la potencialidad de reidentificación que puede inferirse, es decir que en un caso concreto, si sabemos que Juana Rivero de 19 años, oriunda de Artigas está en la base de datos de la institución de salud, padece de esclerosis o listeria.

Además, se puede combinar con técnicas de generalización para obtener mayores garantías de privacidad. Es posible que haya que aplicar otras técnicas para garantizar que un registro no sirva para identificar a una persona.

4.2.2. Diversidad l y Proximidad t

La diversidad l extiende el anonimato k para garantizar que ya no se puedan realizar ataques por inferencia deterministas. Para ello, se debe prevenir que en cada clase de equivalencia, todos los atributos tienen al menos l valores diferentes.

Uno de los objetivos consiste en limitar la ocurrencia de clases de equivalencia que tengan una variabilidad de atributos escasa. Por lo que, si quien quiere reidentificar al titular del dato posee conocimientos previos sobre este, siempre estará sometido a un grado significativo de incertidumbre.

La diversidad l es útil para proteger los datos ante ataques por inferencia, esto es cuando se efectúa el análisis de datos con el fin de obtener ilegítimamente conocimientos sobre un tema o una base de datos, para determinar lo que debe ser protegido en un mayor nivel de seguridad siempre que los valores de los atributos estén bien distribuidos.

La proximidad t es un perfeccionamiento de la diversidad l. Consiste en crear clases equivalentes que se parezcan a la distribución inicial de los atributos en la tabla. Esta técnica es útil cuando haya que conservar los datos lo más próximo posible a los originales. Para ello, se añade una nueva restricción a la clase de equivalencia: no basta con que existan al menos l valores diferentes en cada clase de equivalencia, sino que, además, cada valor debe representarse tantas veces como sea necesario a fin de reflejar la distribución inicial de cada atributo.

Igual que ocurre con el anonimato k, la diversidad l y la proximidad t garantizan que los registros relativos a unapersona no se puedan distinguir o destacar de las otras personas en la base de datos.

Tanto la diversidad l y la proximidad t, en relación al anonimato k, no se pueden llevar a cabo ataques por inferencia contra una base de datos con diversidad l o proximidad t con un cien por ciento de confianza.

Con respecto a la diversidad l se producen insuficiencias que se trata de ejemplificar en la Tabla 3 que sigue:

Año de nacimiento	Sexo	Código Postal	Departamento	Diagnóstico
1950	F	11*	Montevideo	Cáncer
1963	F	11*	Montevideo	Cardiopatía
1967	F	11*	Montevideo	Neuromialgia
1963	F	11*	Montevideo	Ictus-Acv
1955	F	11*	Montevideo	Ictus-Acv
1952	F	11*	Montevideo	Cáncer
1963	F	11*	Montevideo	Cardiopatía
1939	F	11*	Montevideo	Neuromialgia
1970	F	11*	Montevideo	Cáncer
1970	F	11*	Montevideo	Colesterol
1963	F	11*	Montevideo	Cardiopatía
1963	F	11*	Montevideo	Ictus-Acv

Tabla 3: con diversidad l cuyos valores para el atributo "Diagnóstico" no se han distribuido de manera uniforme.

Sin embargo, si se está al tanto que una de las personas nació en 1963 y que aparece en esta tabla, se podría deducir que, muy probablemente, sufrió un ICTUS-ACV como refleja la Tabla 4:

Apellido	Año de nacimiento	Sexo
Jimenez	1950	F
Rosas	1963	F
Pérez	1967	F
Roque	1963	F
Lima	1955	F
Gonzalez	1952	F
Alonso	1963	F
Maneiro	1939	F
Castellano	1970	F
Reta	1970	F
Berreta	1963	F
Gross	1963	F

Tabla 4: Si un atacante supiera que estas personas están en la tabla 3, podría inferir que sufrieron un ICTUS-ACV.

Capítulo 5. Riesgos en el uso de la anonimización

Cuando los responsables de una base de datos o encargado de tratamiento utilizan la anonimización de los datos personales deben recodar que existen riesgos asociados al uso de los datos que han sido anonimizados.

A continuación se analizan los principales riesgos del uso de la anonimización que pueden tener como consecuencia la reidentificación del titular del dato.

5.1. Confusión en la metodología.

Uno de los principales riesgos es confundir y pensar que la seudonimización es lo mismo que la anonimización, lo que no es así.

Los datos seudonimizados se utilizan para ocultar identidades y casi siempre queda un rastro entre el seudónimo y la identidad del titular del dato que corresponde, de manera que permite establecer quién es la persona y vincularlo con otros conjuntos de datos. Este tipo de datos entra dentro del ámbito de aplicación del régimen jurídico de la protección de datos.

Ejemplos de seudonimización:

a) cuando se ocultan las identidades para uso estadístico, científico o de investigación,

b) cuando se cifran los datos y se utiliza la clave para descifrarlos.

Uno de los casos más conocidos es el de los concursos literarios en los que se solicita que el o los autores se identifiquen con un seudónimo y a su vez proporcionen sus datos por separado, para garantizar la imparcialidad del tribunal o del evaluador.

5.2. Confusión en la aplicación normativa.

Otro riesgo que se corre es pensar que los datos que fueron correctamente anonimizados quedan fuera del ámbito de aplicación de la normativa sobre protección de datos, por ende se priva a las personas de cualquier tipo de protección.

Es necesario hacer dos aclaraciones en este sentido: en primer lugar, pese a que los datos hayan sido correctamente anonimizados de manera que cumplan con todas las exigencias técnicas y legales, ante la menor inconsistencia frente a técnicas de inferencia o cuando se presente una posibilidad de reidentificación siempre han de operar las garantías de la Ley de Protección de Datos.

En segundo lugar, el hecho de que en determinados supuestos se excluya la aplicación de la normativa de protección de datos, no significa que otros actos legislativos no sean aplicables al uso de esos datos, como por ejemplo la Ley de Acceso a la Información Pública.

Capítulo 6. Reglas mínimas para la difusión de información judicial en Internet

Especial atención requiere la información que se encuentra en el ámbito del Poder Judicial, ya que la jurisprudencia se ha comenzado a difundir a través de Internet, lo que genera la preocupación en determinar qué grado de protección deben tener los datos personales que se encuentran en las sentencias y fallos judiciales.

Cabe señalar que en Uruguay según lo determina el Código General del Proceso (Art. 7° de la Ley N° 15.982) todo proceso será de conocimiento público, salvo que expresamente la ley disponga lo contrario o el tribunal así lo decida por razones de seguridad, de moral o en protección de la personalidad de alguna de las partes. Es por tal motivo que cobra especial relevancia lograr un balance adecuado entre la publicidad debida y la protección de los datos de carácter personal.

Durante el Seminario “Internet y Sistema Judicial” realizado los días 8 y 9 de julio de 2003 en la ciudad de Heredia, Costa Rica, fueron aprobadas ciertas recomendaciones a ser aplicadas en el ámbito de los Poderes Judiciales. Las denominadas “Reglas de Heredia” fueron aprobadas durante el seminario mencionado, en el cual participaron representantes de los Poderes Judiciales, organizaciones de la sociedad civil y académicos de Argentina, Brasil, Canadá, Colombia, Costa Rica, Ecuador, El Salvador, México, República Dominicana y Uruguay.

Fueron aprobadas diez reglas mínimas que pretenden convertirse en la mejor alternativa o punto de partida para lograr un equilibrio entre transparencia, acceso a la información pública y derechos de privacidad e intimidad.

6.1. Reglas de Heredia

A continuación se detalla el contenido de las diez *Reglas de Heredia*.

Durante la discusión sobre el contenido de estas se expusieron algunos argumentos para la publicación de la jurisprudencia en los sitios de Internet de los diversos Poderes Judiciales. Se argumenta que la finalidad de difundir y publicar las sentencias radica en la necesidad de facilitar el trabajo de las profesiones jurídicas proporcionándoles con celeridad la información en forma completa y actualizada, informar a todo interesado en el proceso, hacer públicas en forma más rápida las nuevas resoluciones adoptadas, contribuir a la coherencia de la jurisprudencia, entre otros argumentos referenciados, según recomendación N° R(95) del Comité de Ministros de la Unión Europea.

Las reglas 1 y 2 refieren a la finalidad:

Regla 1. La finalidad de la difusión en Internet de las sentencias y resoluciones judiciales será:

a) el conocimiento de la información jurisprudencial y la garantía de igualdad ante la ley

b) para procurar alcanzar la transparencia de la administración de justicia.

Regla 2. La finalidad de la difusión en Internet de la información procesal será garantizar el inmediato acceso de las partes o quienes tengan un interés legítimo en la causa, a sus movimientos, citaciones o notificaciones.

La regla 3 refiere al derecho de oposición del interesado y establece:

Regla 3. Se reconocerá al interesado el derecho a oponerse, previa petición y sin gastos, en cualquier momento y por razones legítimas propias de su situación particular, a que los datos que le conciernan sean objeto de difusión, salvo cuando la legislación nacional disponga otra cosa. En caso de determinarse, de oficio o a petición de parte, qué datos de personas físicas o jurídicas son ilegítimamente difundidos, deberá ser efectuada la exclusión o rectificación correspondiente.

Ejemplo de esta regla es utilizar inicialización en el nombre de las personas que han cometido delitos y son primarios.

La regla 4 hace referencia a la adecuación al fin:

Regla 4. En cada caso los motores de búsqueda se ajustarán al alcance y finalidades con que se difunde la información judicial.

De la regla 5 a la 9 refieren al equilibrio necesario entre transparencia y privacidad:

Regla 5. Prevalecen los derechos de privacidad e intimidad, cuando se traten datos personales que se refieran a niños, niñas, adolescentes (menores) o incapaces; o asuntos familiares; o que revelen el origen racial o étnico, las opiniones políticas, las convicciones religiosas o filosóficas, la pertenencia a sindicatos; así como el tratamiento de los datos relativos a la salud o a la sexualidad; o víctimas de violencia sexual o domestica; o cuando se trate de datos sensibles o de publicación restringida según cada legislación nacional aplicable o hayan sido así considerados en la jurisprudencia emanada de los órganos encargados de la tutela jurisdiccional de los derechos fundamentales.

En este caso se considera conveniente que los datos personales de las partes, coadyuvantes, adherentes, terceros y testigos intervinientes, sean suprimidos, anonimizados o inicializados, salvo que el interesado expresamente lo solicite y ello sea pertinente de acuerdo a la legislación.

Regla 6. Prevalece la transparencia y el derecho de acceso a la información pública cuando la persona concernida ha alcanzado voluntariamente el carácter de pública y el proceso esté relacionado con las razones de su notoriedad. Sin embargo, se considerarán excluidas las cuestiones de familia o aquellas en las que exista una protección legal específica.

En estos casos podrán mantenerse los nombres de las partes en la difusión de la información judicial, pero se evitarán los domicilios u otros datos identificatorios.

Regla 7. En todos los demás casos se buscará un equilibro que garantice ambos derechos. Este equilibrio podrá instrumentarse:

a) en las bases de datos de sentencias, utilizando motores de búsqueda capaces de ignorar nombres y datos personales

b) en las bases de datos de información procesal, utilizando como criterio de búsqueda e identificación el número único del caso

Se evitará presentar esta información en forma de listas ordenadas por otro criterio que no sea el número de identificación del

proceso o la resolución, o bien por un descriptor temático.

Regla 8. El tratamiento de datos relativos a infracciones, condenas penales o medidas de seguridad, sólo podrá efectuarse bajo el control de la autoridad pública. Sólo podrá llevarse un registro completo de condenas penales bajo el control de los poderes públicos.

Regla 9. Los jueces cuando redacten sus sentencias u otras resoluciones y actuaciones, harán sus mejores esfuerzos para evitar mencionar hechos inconducentes o relativos a terceros, buscarán solo mencionar aquellos hechos y datos personales estrictamente necesarios para los fundamentos de su decisión, tratando de no invadir la esfera íntima de las personas mencionadas.

Se exceptúa de la regla anterior la posibilidad de consignar algunos datos necesarios para fines meramente estadísticos, siempre que sean respetadas las reglas sobre privacidad contenidas en esta declaración. Igualmente se recomienda evitar los detalles que puedan perjudicar a personas jurídicas (morales) o dar excesivos detalles sobre los modus operandi que puedan incentivar algunos delitos. Esta regla se aplica en lo pertinente a los edictos judiciales.

Y por último la Regla 10:

Regla 10. En la celebración de convenios con editoriales jurídicas deberán ser observadas las reglas precedentes.

6.2. Potencialidad de su aplicación en la administración.

Como punto final corresponde indicar que las *Reglas de Heredia* son simples recomendaciones que se circunscriben a la difusión en Internet o en cualquier otro formato electrónico de sentencias judiciales e información procesal. Por tal motivo el acceso a documentos en formato papel en las oficinas queda fuera de su ámbito de aplicación.

Como se detalló anteriormente se trata de reglas mínimas, es decir, su aplicación no debe desobedecer a la normativa que ofrezca mayores garantías en materia de la protección de los derechos de intimidad y privacidad. De esta manera, tanto las autoridades judiciales como los particulares, organizaciones o las empresas privadas que difundan información judicial en Internet deberán dar cumplimiento a la normativa específica en la materia a nivel nacional y aplicar en todos los casos aquellos procedimientos más rigurosos de protección.

Los proveedores comerciales de jurisprudencia o información judicial deberán por extensión velar por la aplicación de las mencionadas Reglas, si bien están dirigidas a los sitios en Internet de los Poderes Judiciales.

Capítulo 7. Recomendaciones.

Como regla general, es necesario considerar que los datos personales se deben recoger y tratar de acuerdo con la legislación vigente sobre su conservación en un formato identificable.

En consecuencia, el proceso de anonimización, es un caso particular de “tratamiento posterior” a la recolección de los datos personales. Por lo tanto, este tipo de tratamiento debe cumplir con el principio de finalidad para el que fueron recolectados (art. 8º de la Ley Nº 18.331).

En cuanto a la anonimización y disociación se debe estar al caso concreto, por lo que algunas veces para no reidentificar al titular del dato no es suficiente solo utilizar esta técnica sino que se deberá utilizar varias al mismo tiempo.

De igual forma, deben adoptarse buenas prácticas para minimizar los riegos que puedan aparecer con la anonimización o disociación, y quienes realicen el tratamiento de los datos, (responsables), deben concentrar su atención en los medios que serían necesarios para la reidentificación, principalmente en lo que atañe a los conocimientos asociados al uso de dichos medios, a la valoración de la probabilidad y gravedad del uso, a la tecnología disponible al momento del tratamiento (Reglamento General de Datos Personales considerando 23), así como en el costo que le puede ocasionar.

Otro punto a tener en cuenta es que el riesgo de la reidentificación con el tiempo puede aumentar o cambiar, como también puede suceder con la tecnología, por lo que se recomienda que si se procede a la regulación, esta debe considerar la presencia en su articulado del principio de neutralidad tecnológica, es decir que debe ser neutro en cuanto a la tecnología que se regula y pensar en la posible evolución que esta pueda experimentar.

Se entiende que una anonimización realmente resulta eficaz cuando imposibilita distinguir a una persona en un conjunto de datos, o inferir cualquier tipo de información a partir de ese conjunto.

Como principio, para garantizar que ya no se puede identificar al titular del dato, no basta con eliminar los elementos que pueden servir para identificarlo directamente. Por lo tanto se debe tomar las medidas necesarias adicionales, (por ejemplo realizar análisis periódicos de riesgo de reidentificación y cruzamientos de datos), para evitar que se produzca la identificación del titular, la que va a depender del contexto y de los fines del tratamiento que va a ser objeto.

7.1. Recomendaciones para el ámbito público.

Cuando el Estado tome la decisión de que ciertos datos estén disponibles en forma libre y sin restricciones a todos los individuos, en función de la filosofía y práctica que guía a los datos abiertos, debe atender al hecho de que los datos susceptibles de apertura son toda aquella información que posea y que no se encuentre en las categorías que se detallan a continuación:

1. Datos o información secretos, reservados o confidenciales de acuerdo con la Ley Nº 18.381, de 17 de octubre de 2008, artículos 9º y siguientes.

2. Cuando se crea un registro de usuarios, protegiendo su identidad, manteniendo el anonimato y privacidad conforme con las disposiciones normativas vigentes en la Ley Nº 19.172, de 20 de diciembre de 2013, artículo 28 B.

3. Datos personales de acuerdo con la Ley Nº 18.331, de 11 de agosto de 2008, artículo 4º literal D.

Como excepción se puede realizar la apertura de los datos que contengan información personal, siempre que:

a) se cuente con el consentimiento del titular del dato en forma previa, pudiendo abrirlos directamente sin disociación o anonimización,

b) se cumpla con lo establecido en los artículos: 9º en sus diversos literales, entre los que se hace referencia a los listados (es necesario que estos listados existan), para que los datos como nombre, apellido, entre otros, no necesiten el consentimiento, 9º bis, 17 y 23 de la Ley de Protección de Datos y Dictamen Nº 26, de 8 de agosto de 2013, de la Unidad Reguladora y de Control de Datos Personales (URCDP), relativo a la interpretación del alcance del concepto “listado”.

c) si no se cumple alguno de los literales anteriores, los datos estén disociados o anonimizados.

Existen ciertos casos en los que la Administración se encuentra ante información que, por sus características, y para dar cumplimiento al deber de transparencia impuesto por la Ley de Acceso a la Información Pública, debe ser publicada en el sitio web del organismo (por ejemplo Curriculum Vitae de una autoridad).

Sin embargo, pueden existir documentos en los que coexista información que pueda ser conocida e información que debe denegarse en virtud de causa legal, confiriéndose acceso a la primera y no a la segunda en función del principio de divisibilidad. Ejemplo: un titular del dato personal le entrega a la Administración información para la realización de un trámite y solicita que uno o más datos personales sean clasificados como confidenciales. Cada vez que se va a entregar información a un tercero y aparezca o aparezcan los datos personales clasificados como confidenciales deberá aplicarse el principio de divisibilidad, o realizar una versión pública del expediente.

Para efectuar una versión pública se pueden realizar los siguientes procedimientos:

a) sustituir los nombres de las personas por letras (proceso conocido como inicialización), también los números de expedientes, direcciones, entre otros, que puedan identificar directa o indirectamente a las personas.

b) tachar la información identificatoria para que no sea visible por terceros en la publicación.

4. Los organismos que integran el Sistema Estadístico Nacional deben cumplir con el principio de secreto estadístico que obliga a tratar los datos individuales proporcionados por la fuente de información con la más absoluta confidencialidad, de forma tal de no revelar la identificación de dichas fuentes según la Ley Nº 16.616, de 20 de octubre de 1994, artículo 3º.

7.2. Recomendaciones para el ámbito privado.

Cuando una empresa o particular quiera publicar información que contenga datos personales deberá estrictamente haber informado la finalidad, contar con el consentimiento del titular del dato o verificar alguna de las excepciones al consentimiento que plantea la Ley N° 18.331 antes mencionada.

Siguiendo el esquema de las diversas etapas del proceso de anonimización, cobra relevancia asignar roles con diferentes perfiles a nivel interno para lograr que solo se acceda a un mínimo de información imprescindible para realizar su tarea, es decir no todos acceden a toda la información en todo momento.

Si se desea mantener los datos personales para fines estadísticos, científicos o históricos se deberá solicitar la autorización a la Unidad Reguladora y de Control de Datos Personales (URCDP) según Decreto Nº 414/009, de 31 de agosto de 2009, artículos 37 y 38.

Capítulo 8. Consideraciones finales.

Es importante conocer y tener en cuenta que la anonimización, disociación, seudonimización, cifrado, y las demás técnicas antes mencionadas, representan diferentes grados de ruptura del vínculo entre el dato y la identificación de su titular. Los responsables o encargados del tratamiento de los datos tienen que estar informados que un conjunto o paquete de datos anonimizados o disociados pueden tener riesgos residuales para los interesados y adoptar las medidas para prevenirlos. Esto lleva a que la anonimización o la disociación no sea una forma estática de tratamiento de los datos, ni como tampoco un procedimiento ocasional, sino que se tiene que tener presente y evaluar los riesgos que se pueden producir por los responsables y encargados de tratamiento, para que no se pueda reidentificar al titular del dato. Se puede decir que es una forma viva del dato, deben ser conscientes de que un conjunto de datos anonimizado puede entrañar todavía riesgos residuales para los interesados.

Si se opta en el tratamiento de los datos personales, por el resto de las técnicas mencionadas anteriormente (cifrado, seudonimización entre otros), aquel no quedará excluido de la Ley de Protección de Datos Personales.

Las técnicas de anonimización pueden aportar garantías a la protección de datos personales siempre que su aplicación se diseñe en forma adecuada, por lo que se deben tener en cuenta las etapas del proceso de anonimización (preanonimización, anonimización y control).

Como último punto se tiene que recordar que deben aplicarse las recomendaciones mencionadas en el Capítulo 7 para garantizar los derechos de las personas en cuanto a su identificación.

Igualmente se debe emplear, en lo que sean aplicables, las reglas mínimas ("Reglas de Heredia"), para la difusión de la información judicial en Internet, como una buena práctica y orientación a quienes se encargan de la difusión de las sentencias.

En suma, la anonimización no debe verse como un procedimiento temporal, todo lo contrario, y los responsables de la tarea deben evaluar en diferentes oportunidades si existe riesgo o no de reidentificación del titular de los datos, para en su caso subsanar la situación.